



Poste de Post-Doctorat

Détection de la fraude dans des documents d'entreprise

Le LabCom IDEAS lance un appel à candidatures pour un poste de post-doctorant en informatique dans le domaine de la détection de la fraude sans connaissance a priori dans des documents d'entreprise.

Durée : 18 mois

Date d'embauche souhaitée : 1^{er} février 2021 (pouvant être modifiée en fonction de la situation sanitaire)

Salaire : environ 2100 € net / mois

Lieu de travail : LabCom IDEAS, dans les locaux du laboratoire L3i à La Rochelle, France

Spécialités : Informatique / Traitement et analyse d'images / Traitement du signal / Image forensics

Description du LabCom :

Les travaux menés par le candidat s'inscriront dans le cadre du LabCom IDEAS, cofinancé par l'Agence Nationale de la Recherche (ANR) et la région Nouvelle-Aquitaine, et regroupant l'entreprise Yooz et le laboratoire L3i. Ce LabCom a pour objectif d'imaginer, inventer, concevoir, développer, optimiser et entraîner les meilleurs algorithmes automatiques de traitement des documents d'entreprise pour offrir un service d'intelligence artificielle capable de comprendre un maximum de document d'entreprise.

Le post-doctorant sera basé au sein du LabCom, localisé dans les locaux du Laboratoire Informatique, Image et Interaction (L3i), à La Rochelle.

Le laboratoire L3i, EA 2118 créé en 1993, représente la seule composante de recherche du domaine STIC à l'Université de la Rochelle associant les chercheurs de l'IUT de la Rochelle, et du Faculté des Sciences et Technologies en informatique.

Le large déploiement des technologies numériques et la multiplicité des processus d'acquisition et de diffusion de l'information engendrent un développement rapide et diversifié des modes de production et de consommation de contenus numériques, ainsi qu'une croissance exponentielle de la volumétrie des données. Par ailleurs, l'avènement des dispositifs nomades interactifs augmente encore plus les problématiques de positionnement de l'utilisateur dans la gestion et la navigation au sein de contenus numériques.

Il s'agit, pour le L3i, de mettre en synergie les compétences établies dans le laboratoire afin d'aborder la problématique de la valorisation des contenus numériques sous un angle systémique. Cela revient, en particulier, à une exploitation croisée des compétences en matière d'applications interactives, d'indexation par le contenu, et de représentation de connaissances. Le laboratoire se structure autour de trois thématiques scientifiques (Ingénierie des connaissances, Analyse et gestion de contenus, Interactivité et dynamique des

systèmes), toutes centrées sur la problématique de la gestion interactive et intelligente des contenus numériques.

Yooz, partenaire industriel du Labcom, est fournisseur d'un service Cloud d'automatisation des processus d'achat et de paiement. Yooz intègre des technologies d'Intelligence Artificielle pour automatiser les processus et le traitement des documents impliqués dans ces processus. Le service Yooz est utilisé quotidiennement par près de 3000 utilisateurs.

Le travail de recherche et développement mené au sein du LabCom s'articule autour de 3 grands axes :

- Classification de documents
- Fouille de documents
- Détection de fraude documentaire

Description du poste :

Le travail du post-doc recruté s'inscrira dans le cadre de l'axe "Détection de fraude documentaire". Il s'agit de développer des méthodes de détection des modifications frauduleuses (*image forensics*) sans connaissance *a priori* dans les images de document.

Actuellement, il n'existe aucune solution fiable pour protéger les entreprises des fraudes documentaires. Malgré les besoins importants, la détection des fraudes automatique dans les documents est peu étudiée. La plupart des travaux sur la détection des documents falsifiés concernent des indices graphiques, comme la différence d'inclinaison, de taille, d'alignement ou de bruit d'un caractère par rapport aux autres, la différence de police ou d'espacements des caractères au sein d'un mot [Bertrand15], le décalage d'une ligne par rapport aux marges ou encore l'inclinaison différente des lignes les unes par rapport aux autres. Dans ces travaux, l'hypothèse de départ est que les fraudeurs doivent souvent modifier les éléments du document dans la précipitation et que la modification n'est donc pas toujours parfaite, ce qui permet de la détecter. De plus, ce type de méthode nécessite une connaissance *a priori* sur le type de fraude réalisé et l'erreur commis par le fraudeur ainsi que des propriétés du document.

Ce travail de post-doctorat s'appuiera sur un état de l'art détaillé des approches existantes, pour en identifier les limites, et proposer des approches innovantes qui permettent de contourner les inconvénients mentionnés ci-dessus. Dans nos travaux précédents, nous avons développé plusieurs méthodes de détection de régions fraudées dans les images de document, *e.g.* [Cruz17] qui propose une méthode sans connaissance *a priori* sur la fraude ni sur les propriétés du document en se basant sur des *local binary patterns* (LBP). Le travail de ce post-doc s'inscrit dans la même ligne. Plus précisément, le travail visera à développer une méthode de détection de fraudes en s'appuyant sur la transformée en cosinus discrets (DCT). Cette approche de vérification d'une image à partir de la DCT est répandue dans les images de scènes naturelles car elle présente une excellente propriété de regroupement de l'énergie et permet de faire porter l'information par seulement quelques coefficients. Cependant, cette approche n'a pas été étudiée pour des images de documents. Le travail consistera donc à

apprendre (via des approches de type *deep learning*) les variations d'énergie des coefficients DCT correspondent à un acte de fraude ou non.

Profil recherché :

Le candidat, titulaire d'un doctorat dans les domaines de l'informatique, du génie informatique et traitement du signal, ou des mathématiques appliquées, devra justifier d'une expérience de recherche dans le domaine de traitement et d'analyse d'image ou du signal.

Les compétences du candidat incluront :

- Maîtrise nécessaire d'un ou plusieurs langages de programmation (Java, Python, C/C++...)
- Très bonnes aptitudes au travail en équipe, une connaissance des méthodes Agile serait un plus (le travail sera mené à la fois en lien avec les chercheurs du laboratoire L3i et le service R&D de l'entreprise Yooz)
- Bonne aptitude à la rédaction d'articles scientifiques et maîtrise de l'anglais écrit et parlé

Pour postuler :

Les candidats à ce poste devront envoyer un CV et une lettre de motivation (les noms et coordonnées de références seraient un plus) à :

- [petra.gomez \[chez\] univ-lr.fr](mailto:petra.gomez@univ-lr.fr)
- [mickael.coustaty \[chez\] univ-lr.fr](mailto:mickael.coustaty@univ-lr.fr)
- [nicolas.sidere \[chez\] univ-lr.fr](mailto:nicolas.sidere@univ-lr.fr)
- [Vincent.PoulainDAndecy \[chez\] getyooz.com](mailto:Vincent.PoulainDAndecy@getyooz.com)
- [Saddok.Kebairi \[chez\] getyooz.com](mailto:Saddok.Kebairi@getyooz.com)

Les candidatures seront étudiées au fil de l'eau et donc il n'y a pas de date limite de candidature.

Références :

[Bertrand15] R. Bertrand, O. Ramos Terrades, P. Gomez-Krämer, P. Franco, J.-M. Ogier, A Conditional Random Field model for font forgery detection, ICDAR 2015, p. 576-580.

[Cruz17] F. Cruz, N. Sidere, M. Coustaty, V. Poulain D'Andecy, J.-M. Ogier, Local Binary Patterns for Document Forgery Detection. ICDAR 2017, p. 1223-1228.